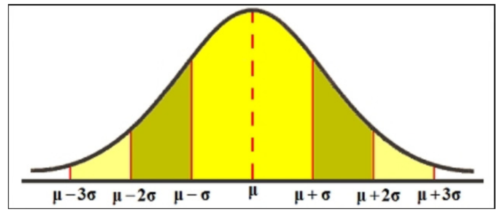
History of Statistics 2. **Origin of the Normal Curve – Abraham DeMoivre (1667- 1754)**

The normal curve is perhaps the most important probability graph in all of statistics.

Its formula is shown here with a familiar picture. The "*e*" in the formula is the irrational number we use as the base for natural logarithms. μ and σ are the mean and standard deviation of the curve.

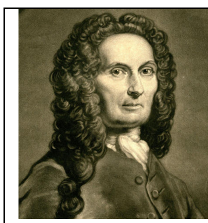
$$y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



The person who first derived the formula, **Abraham DeMoivre** (1667- 1754), was solving a gambling problem whose solution depended on finding the *sum of the terms of a binomial distribution.* Later work, especially by **Gauss** about 1800, used the normal distribution to describe the pattern of random measurement error in observational data. Neither man used the name "normal curve." That expression did not appear until the 1870s.

The normal curve formula appears in mathematics as a *limiting* case of what would happen if you had an infinite number of data points. To *prove* mathematically that some theoretical distribution is actually normal you have to be familiar with the idea of limits, with adding up more and more smaller and smaller terms. This process is a fundamental component of calculus. So it's not surprising that the formula first appeared at the same time the basic ideas of calculus were being developed in England and Europe in the late 17^{th} and early 18^{th} centuries.

The normal curve formula first appeared in a paper by DeMoivre in 1733. He lived in England, having left France when he was about 20 years old. Many French Protestants, the Huguenots, left France when the King canceled the Edict of Nantes which had given them civil rights. In England DeMoivre became a good friend of Isaac Newton and other prominent mathematicians.



Abraham DeMoivre

He wrote the 1733 paper in Latin, and in 1738 he translated it himself into English for the 2^{nd} edition of his book, ***The Doctrine of Chances***, one of the first textbooks on probability. (The first edition had been published in 1718.)

In DeMoivre's work the normal curve formula did not *look* like it does now, in particular because there was no notation then for *e.* and there was no general sense of standard deviation, which is represented by σ in today's equation.

Why did DeMoivre do it? What problem was he working on?

Here's how his paper starts. (You can read it even though it is in old style English.)

You can see that he is dealing with "Problems of Chance," that he wants to see how likely it is that an "experiment" will produce a given outcome. Note that he credits the Bernoulli brothers with prior work – but they just didn't do quite enough.

The core problem for DeMoivre is to find the sum of "several" terms in a binomial expansion. He wanted a shortcut because the problem was "so laborious."

A Method of approximating the Sum of the Terms of the Binomial $\overline{a+b}\backslash^n$ expanded to a Series from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments.

Altho' the Solution of Problems of Chance often require that several Terms of the Binomial $\overline{a+b}\backslash^n$ be added together, nevertheless in very high Powers the thing appears so laborious, and of so great a difficulty, that few people have undertaken that Task; for besides *James* and *Nicolas Bernoulli*, two great Mathematicians, I know of no body that has attempted it; in which, tho' they have shewn very great skill, and have the praise which is due to their Industry, yet some things were farther required; for what they have done is not so much an Approximation as the determining very wide limits, within which they demonstrated that the Sum of the Terms was contained.

DeMoivre wanted to avoid having to add up all these coefficients. He needed to describe the general shape of the distribution of the values on a line of coefficients without having to compute each one. We can see what happens with a few graphs.

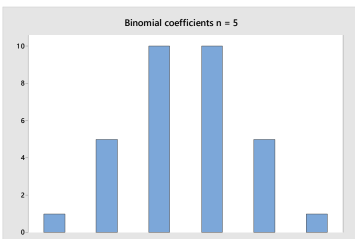
A clear example of this problem can be seen in Table 1. Imagine that you want to find the sum of several terms in one line, say the middle two terms in line for n = 5.

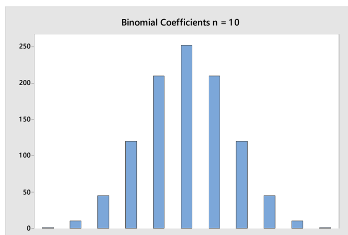
We quickly see that 10 + 10 = 20.

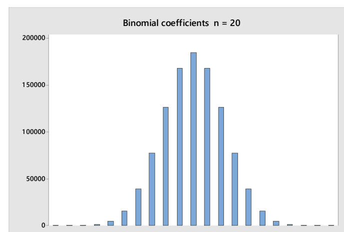
But what if you want to find the sum of the middle 10 terms in the line where n = 100? A problem like this could easily come up in a game of chance. This is what he meant by "laborious."

Table 1			
n	Expansion of $(a+b)^n$	Coefficients	Sum 2^n
1	a+b	1 1	2
2	$a^2+2ab+b^2$	1 2 1	4
3	$a^3+3a^2b+3ab^2+b^3$	1 3 3 1	8
4		1 4 6 4 1	16
5		1 5 10 10 5 1	32
	etc.	etc.	etc.

DeMoivre wanted to avoid having to add up all these coefficients. A solution is to describe the general shape of the distribution of the values on a line of coefficients without having to compute each one. We can see what happens with a few graphs.







If you can replace the binomial coefficient graph that is made up of lots of bars by a smooth curve, then instead of having to add lots of individual numbers you can just find the area under the curve, which is exactly one of calculus's strengths. You can see that as *n* increases the graphs look more and more like a bell-curve.

DeMoivre began by considering the expansion of $(1+1)^n$. This expression comes up naturally in analyzing a coin toss with equal probabilities of heads and tails. He focused on the *ratio* of the *middle* term to the sum of all the coefficients, 2^n.

Here are a couple of concrete examples.

For $n = 2$, you can see from Table 1 that the ratio of the middle term to the sum is $\frac{2}{4} = .5$. If you look at $\frac{(1+1)^2}{2^2} = \frac{1+2+1}{4}$, you can see again where $\frac{2}{4}$ comes from.

For $n = 3$ the ratio will be $\frac{3}{8} = .375$. (When *n* is odd we can take the "middle term" to be the value of either of the two middle terms.)

DeMoivre wanted to know what happens to the ratio when n gets very large. It is informative to see how a 17th century mathematician has to deal with such a problem.

Here's how he gets into the problem:

> I. It is now a dozen years or more since I had found what follows; If the Binomial $1+1$ be raised to a very high Power denoted by n, the ratio which the middle Term has to the Sum of all the Terms, that is, to 2^n, may he expressed by the Fraction $\frac{2A\times\overline{n-1}\backslash^n}{n^n\times\sqrt{n-1}}$, wherein A represents the number of which the Hyperbolic Logarithm is $\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}$, &c. but because the Quantity $\frac{\overline{n-1}\backslash^n}{n^n}$ or $\overline{1-\frac{1}{n}}|^n$ is very nearly given when n is a high Power, which is not difficult to prove, it follows that, in an infinite Power, that Quantity will he absolutely given, and represent the number of which the Hyperbolic Logarithm is -1;

To reprise:

He had "a dozen years or more" ago found that the ratio of the middle term to the sum can be expressed, using modern parentheses, as $\frac{2A(n-1)^n}{n^n\sqrt{n-1}}$. In these notes, we refer to the ratio as R.

He wants to determine the value of R for any given *n*. Once he gets that, he knows the height of the curve at the middle and could then build the rest from that. He used some results he had found earlier about limits, some earlier work by James Bernoulli, and work that James Stirling did at just the right time to help.

First DeMoivre rewrites his fraction as $\frac{2A}{\sqrt{n-1}}\left(\frac{n-1}{n}\right)^n$ or $\frac{2A}{\sqrt{n-1}}\left(1-\frac{1}{n}\right)^n$. He wants to find the limiting value of this expression when *n* approaches infinity. Many mathematicians of that time were working on problems that involved expanding binomial expressions.

DeMoivre is able to use earlier work by James Bernoulli to say that the limit for $\left(1-\frac{1}{n}\right)^n$ is "the number whose hyperbolic logarithm is -1." Today we say "natural" logarithm instead of hyperbolic logarithm, and we would therefore write $\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = e^{-1}$.

So, for large values of *n* we can replace $\left(1-\frac{1}{n}\right)^n$ by e^{-1}. That gets us to the ratio as R = $\frac{2Ae^{-1}}{\sqrt{n-1}}$.

He also states that he has shown in his own earlier work that *A* is the number whose hyperbolic logarithm is $\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\cdots$

In today's notation we would write $A = e^{\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\cdots}$.

That gets us to R = $\frac{2\left(e^{\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\cdots}\right)(e^{-1})}{\sqrt{n-1}}$ or $\frac{2e^{-1+\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\cdots}}{\sqrt{n-1}}$.

It was easier for DeMoivre to have the exponential expression in the denominator of R.

So we can write R = $\frac{2}{\left(e^{1-\frac{1}{12}+\frac{1}{360}-\frac{1}{1260}+\frac{1}{1680}+\cdots}\right)\sqrt{n-1}}$.

DeMoivre called the messy looking part *B*. So he ends up with R = $\frac{2}{B\sqrt{n-1}}$..

DeMoivre says this is as far as had gotten when he had last worked on the problem. It had been good enough for him to do some approximate calculations but he did not get a "nice" expression for B.

> When I first began that inquiry, I contented myself to determine at large the Value of B, which was done by the addition of some Terms of the above-written Series; but as I pereciv'd that it converged but slowly, and seeing at the same time that what I had done answered my purpose tolerably well, I desisted from proceeding farther, ...

Lucky for him, his friend and colleague, James Stirling, showed the exact value of *B*.

> ... till my worthy and learned Friend Mr. *James Stirling*, who had applied himself after me to that inquiry, found that the Quantity B did denote the Square-root of the Circumference of a Circle whose Radius is Unity, ...

Since the circumference of a circle whose radius is 1 is 2π, we can write DeMoivre's formula for the desired ratio as R = $\frac{2}{\sqrt{2\pi}\sqrt{n-1}}$.

Then he reasoned that for really large values of *n* there's no appreciable difference between $\sqrt{n-1}$ and $\sqrt{n}$. We can therefore write the ratio as R = $\frac{2}{\sqrt{2\pi}\sqrt{n}}$. This starts to look like the modern formula for the normal curve.

We can show the similarity further remembering that DeMoivre is approximating a binomial with $p = ½$. The standard deviation for a binomial distribution is given by

$\sigma = \sqrt{np(1-p)}$ and replace p by ½ to get $\sigma = \sqrt{n\frac{1}{2} \times \frac{1}{2}} = \frac{\sqrt{n}}{2}$

Solving for $\sqrt{n}$ we get $\sqrt{n} = 2\sigma$.

And so finally, R = $\frac{2}{\sqrt{2\pi}\sqrt{n}} = \frac{2}{\sqrt{2\pi}2\sigma} = \frac{1}{\sqrt{2\pi}\sigma}$.

This is exactly the y value of the normal curve at the center where $x = \mu$.

$y = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^0 = \frac{1}{\sqrt{2\pi}\sigma}$.

Some statistical consequences of DeMoivre's work on the normal curve.

The formula itself would be found to have more and more applications, beyond just the approximation of a binomial distribution, especially as a model for the distribution of random measurement errors. Also, equally informative, the relationship $\sigma = \frac{\sqrt{n}}{2}$, shows that *the narrowness of the normal curve, as measured by σ, is proportional to the* ***square-root of the sample size***. This implies, for example, that if you wish to cut a margin of error in half, you will need four times as much information, not twice as much. Another way to say this is that the information you gain in collecting data is proportional, not to the number of pieces of data, but to its square root.

Exercises

1. In connection with Table 1 we showed how to find the ratio of the middle term to the total. What will the ratio be for n = 4? 5? 10? 50? (Really? 50?)

2. Use a calculator to show that in DeMoivre's work $A = e^{\frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680} + \cdots}$ is approximately 1.09. Use a calculator to confirm that $B = e^{1 - \frac{1}{12} + \frac{1}{360} - \frac{1}{1260} + \frac{1}{1680} + \cdots}$ is approximately 2.507 and confirm that B= 2.507 is a good approximation to the square-root of the circumference of a circle whose radius is unity.

Sources:

Stigler, *Seven Pillars of Statistical Wisdom*, Harvard U Press, 2016

Stigler, *The History of Statistics*, Harvard U Press, 1986

Quotations from DeMoivre's Doctrine of Chances. Walker, Helen. *DeMoivre On The Law Of Normal Probability*, www.york.ac.uk/depts/maths/histstat/demoivre.pdf

Image of DeMoivre. www.quo.es/ser-humano/es-posible-calcular-la-fecha-de-la-muerte-de-una-persona

Weisberg, Herbert. *Willful Ignorance: The Measure of Uncertainty*, Wiley 2014